

Appendix A



BsgI restriction endonuclease (RM fusion) coding sequence
3108 nt (recognition sequence: GTGCAG)

ATGAATAGAG	TAGAACTCTAA	AAAAAAATTA	GAACAATTAG	TTCAACAGTT	CGAGAAGTAT
GAAAGTACAT	ATAGCGCTTC	GGATTATAAA	GAGGCAACTT	TAAGATCTAG	TTTTTTAGAT
CCCTTTTTTG	AACTTTTTGG	ATGGGAAATG	CGCCCTGAAA	GAATAACTAA	TCCAGCAGAC
TTAGAAGTGA	TTATAGAAGA	AAGTTTAGAA	ACGGAAAAAT	CTACTAAGTA	TATAGATTAT
GT'TTTTAAAA	TTAATAGAAC	GACTCAGTTT	TTGGTAGAAG	CTAAAAAGCC	AGCTGAAAGT
CTTTC'TAAAA	AAGATCATAT	TTTTCAGGCT	AAAAGTTATG	CATTACTAC	GGAGATTCCA
TTTGTCA'TTT	TAACAAATTT	TAAAGAGTTC	AGATTTTATG	ACGTTTCAAC	TGAACCTTTA
CACAATCAAC	CGGATACAGA	TAAAGTGGAA	GAATATTGTT	TTGATTATAA	AGAATATGTT
CAAAAC'TTTG	ATAAGTTATG	GGAATTATTC	AGCAGAGAAG	CAGTTGCTAA	CAGAAGTTTA
GCCAAG'TTTT	ATGCTAAAAG	AAGAAATATA	GTAGATAGTC	CAGATTTAAT	TTTTAAACTT
AATTATCAAA	TTGATAAAGG	TGCATCATT	CTGGATATAT	CTTTC'TTAAA	AAATTTGAAA
ATATGGAGAA	AATCATTAGC	TGAAAATATC	TTTAATAATA	ATTCACTTAA	TGTTAACGTA
ATTAATGAAG	TAGTTCAGAG	AATATTAGAT	AGACTGATAT	TTATCCGTAT	CATTGAAGAC
AGAAATATTG	AATCTAAAGA	GT'TTTTAAAA	GAAATTGTAG	AAATGCACGA	ACAAGATAAT
TCGATTTTCAG	TGAAAAATGA	ACTAGATAAA	CTATGTATTG	AATTAATAA	GAAATTTAAT
GGGTTAG'TTT	TCCATGACCA	CACATTTGTT	AACGAAGCGT	TGATAGATAA	CGAAATTTTA
ATAGTTATTA	TTGACAATTT	ATATTATCCA	AAGTCTCCTT	ATAACTTTAG	ATTAATTAAG
CCAGAAATTT	TAGGACGAAT	ATTTCGAGCA	TTTTTG'GGTG	AAAAAATGA	AATAATCGAT
GGAAAAATAA	CGTTAGGATT	AAAAGATATT	AATAAAAAAT	CAGGAGGTGT	TTATTATACA
CCTTCATATA	TAGTTGAAAA	AATAGTAGAA	AATACATTAT	CCAAAAAAT	ACATAATGAT
ATTACTATTG	AAAATTTAGA	ACAGATAAAA	ATAGCTGACA	TAGCTTGTGG	TTCAGGAAGC
TTTTTTAATTT	CATCATATAA	ATATTTAATT	GATAAATTTT	AATATATTTA	TTCCAAATGT
TCGGAAGCGG	ATGTTCAAAC	ATTAATTAGT	AATAACTTAG	TATTTATAGA	CAATGGTAAA
TTAATGTTAA	CAATGGAACA	TAAAAAGGGG	ATACTTCAGC	AAAATATTTT	TGGGGTAGAT
ATAGATTCAC	AAGCAATTCA	GGTAGCGAAA	TTAAGTCTTT	ATATAACCAT	GTTAGAAGAA
GGATACAGAG	AAGGTACATT	AAGACCTATA	TTACCAGACT	TAAATGATAA	TATTTAAACAT
GGTAAC'TCAA	TAATAGATAA	TGAAATTTTA	TTTGAAGATG	ATATAAATTA	CGATATTGAT
GCAACATTAC	CATTCGATTG	GGAATATGCT	TTTCCTGATA	TTATAGATAA	CGGAGGTTTT
GATGTAATAT	TAGGCAATCC	ACCCATATATA	AGAATTCAAA	TTTTTGAAGA	GTTATATGGA
AAAGATGTAG	TTAATTATTT	GAAAAA'AAAA	TACGTTTCTG	CCGAAAAAT	TAAC'TTTGAT
ATATATGTCG	TGTTTATAGA	AAAAGCATTG	TCACTCTTGA	ATGACCAGGG	GATATTGGGA
TATATTGTGA	TGAACAAATT	TTTTACTACA	CAATATGGAG	AAAAATTGCG	CGAGTTAATA
ACTTTCACAAA	AATTATTATA	TGAAATCATT	GATTTTGGAA	TTAATGAAAT	ATTTAATAAT
GCTACTACTT	ATACTTGTAT	ATTAATTTTA	GACAAA'ACTA	ATCCAGATGA	AATAATTATT
GAAAGAGTGA	TTGATTTAAA	TACTTGGAAA	GCTGGAGAAT	CTTCAGATCG	GAAAGTGGTA
GATCATACCG	AATTCACTAG	TACTCCTTGG	TATTTATCAA	GCAATACTGA	TGAAGAAATT
TACAAAT'TCT	TTGAAGAAAA	TATGGTTTTTA	CTTGAAACCA	TTAGTGATAG	GGTTTTTTGTT
GGTGTTCAGA	CAGACTGTGA	TCCAGTATAT	ATTTTAGAAG	AAGTTTATGA	AGAAGAAAAAT
TATTTATATT	GTAAGTCAGA	ATATACTACT	GAAGTACACA	AGTTTGAAAA	AGATCATTTA
AAACCATTTT	TAAAAGGTTT	TCTAGATATA	AAGAAATATA	CTTTTTCAAA	TGTTAATAAG
TGGTTACTTT	TCCCTTATAC	CAATTCGGAA	AATACTTCTG	ATTTAATTCC	CGAAACAAC
TACAAACAGT	ATTTCC'CAAG	AACATGGAAA	TACTTAGAGT	CTTGTAAGA	AAGATTAGCA
AAAAGAAAAA	GTATTGAAAG	AGAATTGGAT	ATTAATCCGA	ATTATAATGA	GTGGTATAAA
TATATTTTACA	AAAAGAATCA	CACGAGGATG	GACCAATTAA	AAATAGTATT	TCCTGCGATA
TCGAAGGGTA	GTAGCTTTTG	TTATGATTCT	GATGGAGAGT	ACTATTTTGT	AGGAAGTGGT
GCTGGAGGCG	GTGGTGGAGG	CGCAATAGTC	TTGCCAGATC	AATCTGATTA	TAATTATTTA
TCCTTACTTG	GAATTCTAAA	TTCAGAAGTA	GTTTCATATC	AAATTGTAAG	AAGAGGTTCA
AAACATAAAG	GTTCTTATTA	TGGTGTAGAT	AAAAAGAGAA	TAGAAAATCT	ATATGTGCCA
TTGATTAATG	AGGATAATAA	AAATTTATTT	AGTAATATTT	CAAAAATGGT	AGCTCAAATT
CTTGATGCGT	TTCAAAAAAT	GCATCAAGCA	GGGACAACGG	ATGTTGGTAA	AGAACA'ACTT
CAACAAAGAA	TAAAAATGCT	TAATGCTAGA	ATAAATGAGC	TGGTATATAG	ACTGTATAAT
TTACCAGTAG	AATATAAAGA	ATATATTAAA	AATGCCTTAG	AAAATTAA	